

Automatización del Test de Bechdel-Wallace

José Nahuel Freitas*, Milena Rosenzvit, Stephanie Muller

CONICET – FCEN (UBA)

Recibido: 19 de febrero 2016; aceptado: 12 de marzo 2016

Resumen

Una película pasa el Test de Bechdel-Wallace si cumple que 1) hay al menos dos mujeres en ella que 2) en algún momento hablan entre sí 3) acerca de algo que no es un hombre. Se cree que aproximadamente el 40 % de las películas actuales falla en el Test. A pesar del enorme potencial que tiene esta herramienta para hacer visible la desigualdad de género en el cine, no se ha desarrollado, desde su creación en 1985, una versión automática de la misma. Ese fue el propósito de este trabajo. El test desarrollado permitió, a través del análisis de los guiones del sitio imdb.com, obtener evidencia de que la cantidad de películas que pasan el test es incluso menor a estimaciones previas. Además, este programa permite analizar grandes corpus de guiones en función del género cinematográfico.

Palabras clave: Test de Bechdel-Wallace | Desigualdad de Género | Análisis de textos automáticos | Análisis de guiones

Abstract

A movie is considered to pass the Bechdel-Wallace if it satisfy three conditions: 1) there are at least two women characters that 2) talk to each other at some point 3) about something that is not related to a men. It is believed that about 40% of movies fail to pass the test. In spite of the great potential of this tool to make visible the gender inequality in cinema, since its introduction in 1995, an automatic version of it is still to be developed. That was the purpose of this work. The developed test enabled us, through the analysis of movie scripts obtained from the site imdb.com, obtain evidence that the number of movies that pass the test is even lower than previous estimations. Also, this program allows to analyse large sets of scripts as a function of the cinematographic genre.

Keywords: Bechdel-Wallace test | Gender Inequality | Automatic text analysis | Script analysis

Introducción

El Test de Bechdel-Wallace apareció por primera vez en la tira cómica *The Rule* del cómic *Dikes to Watch Out For*¹ de la historietista estadounidense Alison Bechdel, en 1985. En aquella tira (ver Figura 1), uno de los personajes es invitado al cine y responde que solo acepta ver una película si la misma pasa 3 condiciones: 1) debe haber al menos dos personajes femeninos en la película 2) en algún momento los personajes femeninos deben hablar entre sí, 3) acerca de algo que no sea un hombre. Estas reglas fueron ideadas por Liz Wallace, una amiga de la autora, que dijo haberse inspirado en un fragmento de *A Room of One's Own*, de Virginia Woolf:² “*And I tried to remember any case in the course of my reading where two women are represented as friends. [...] They are now and then mothers and daughters. But almost without exception they are shown in their relation to men. It was strange to think that all the great women of fiction were, until Jane Austen's day, not only seen by the other sex, but seen only in relation to the other sex.*”. Luego de

la publicación, este conjunto de condiciones pasó a ser conocido como El test de Bechdel-Wallace o, como nos referiremos de ahora en adelante, Test de Bechdel.

En el año 2008 se creó el sitio web bechdeltest.com,³ en el cual los usuarios pueden subir el nombre de una película y el resultado de la aplicación del test. Al día de la fecha (14/02/2016) el sitio cuenta con 6341 películas, de las cuales el 57,6 % pasa el test.⁴ Asimismo, muchas de las películas más conocidas y premiadas de la historia, sorprendentemente, no pasan el test. Algunos ejemplos son la trilogía original de *Star Wars*, la saga completa de *El Señor de los Anillos*, *El Padrino*, *Forrest Gump*, y *El Rey León*. Entre las argentinas, podemos citar *El Secreto de sus Ojos*.⁵ El hecho de que una película en particular pase el Test no implica que haya una participación femenina aceptable, o que no tenga contenido sexista. Igualmente, una película puede no pasar el Test de Bechdel y aun así transmitir un mensaje por la igualdad de género. Pero, si analizamos en conjunto un gran número de películas, que sean representativas, por ejemplo, de una época, o poseedoras de un cierto rating, o distintivas de un direc-

* nahuel.freitas@gmail.com



Figura 1. Tira cómica que dio origen al test.

tor o un género cinematográfico, y obtenemos un alto número de ellas que no pasa el test, resulta razonable interpretar los resultados en términos de desigualdad de género. Al ser un test muy poco exigente, permite poner de relieve la poca participación femenina en la ficción cinematográfica. Basados en su relativa sencillez, y en el potencial que tiene este test para mostrar, en términos estadísticos, el lugar de la mujer en el cine, decidimos desarrollar una versión computacional del mismo. Creemos que este desarrollo tiene principalmente dos ventajas respecto al sitio web bechdeltest.com. Por un lado, permite analizar un gran número de películas al mismo tiempo, y comparar los resultados obtenidos en función de distintos parámetros (género, director, fecha de producción, país de procedencia). En segundo lugar, permite hacer el análisis sobre un conjunto de películas que puedan resultar una muestra verdaderamente representativa de un grupo. Por el contrario, la base de datos del sitio web crece con el constante aporte particular de los usuarios, que puede tener diferentes tipos de sesgos. Luego de una exhaustiva búsqueda, no hemos encontrado ningún desarrollo similar, por lo que creemos que puede resultar una innovación significativa.

Métodos

El Test de Bechdel se presta especialmente para hacer una versión computacional puesto que comprende tres simples condiciones que pueden analizarse a partir de los guiones de las películas. No es un test para el que existan distintas interpretaciones acerca de cómo aplicarlo. En distintos sitios web, foros y estudios en general siempre se pretende que el test se aplique de manera totalmente estricta: no influye el largo de los diálogos o la importancia de los mismos, sino simplemente si se cumplen o no esas tres condiciones. En particular, hay una sola enmienda que se suele hacer a la primera condición, y que nosotros decidimos adoptar por cuestiones metodológicas: los personajes femeninos en cuestión deben tener nombre. Así, el test queda resumido de la siguiente manera: una película pasa el test si tiene al menos dos personajes femeninos con nombre que en algún momento hablan entre sí acerca de algo que no es un hombre. Por lo demás, el resto de los personajes pueden ser todos masculinos, o femeninos sin nombres, o tener diálogos totalmente machistas. En términos generales, el programa desarrollado debería: a) identificar los personajes de cada guión y clasificarlos en femeninos y masculinos, estableciendo si hay al menos dos femeninos b) identificar si hay al menos alguna conversación entre dos personajes femeninos y c) para todas las conversaciones entre personajes femeninos, identificar si hay o no una referencia a un hombre. Luego, el programa podría clasificar a las películas según pasen 0, 1, 2 o las 3 condiciones. Para escribir el programa se utilizó el lenguaje Python.⁶ El mismo lenguaje fue utilizado para obtener los guiones y bases de datos automáticamente de los distintos sitios web. El módulo BeautifulSoup fue de gran ayuda al momento de procesar los documentos html.⁷

Indetificación de personajes. El primer paso del programa es convertir cada guión en una lista de intervenciones. Cada intervención consiste en el nombre del personaje y lo que este recita. Luego identifica a cada uno de los nombres de esa lista con el género correspondiente, para lo cual busca ese nombre en listas de nombres femeninos y masculinos en inglés. Estas listas fueron generadas a partir de la base de datos de nombres del censo de Estados Unidos del año 1990.⁸ Sin embargo, hay algunos personajes que nuestro programa no puede identificar, debido a que son nombres especialmente inventados para esa película, o en idiomas distintos al inglés, o bien son nombres que se usan para ambos géneros. No hemos tenido problemas con nombres poco comunes, de-

bido a que las listas generadas son muy completas (1565 nombres de mujeres y 1181 nombres de hombres). Otro tipo de casos se da cuando los personajes están nombrados por su apellido más algún título, por ejemplo: Mr. Brown. Es por ello que las listas de nombres generadas contienen además algunos títulos, como Mr, Ms, Queen, entre otros. A pesar de ello, algunos personajes cuyo título es genéricamente ambiguo (“Professor”) no pueden ser identificados por el programa. Por último, hay casos en los que los personajes no tienen nombres, por ejemplo: “The Waitress”, “The Lawyer”. En estos casos, si bien nuestro programa no puede identificar el género del personaje, esto no constituye un problema para analizar si se cumplen o no las primeras dos condiciones, puesto que hemos adoptado la enmienda de que los personajes femeninos que participen del diálogo en cuestión deben tener nombre. Sin embargo, en el caso de que haya conversaciones entre dos mujeres, esto puede llevar a errores al analizar si se cumple o no la tercera condición.

Identificación de conversaciones. Para identificar si existen diálogos entre mujeres con nombre, el programa recorre todas las intervenciones de un guión e identifica aquellos casos en los que hay dos intervenciones consecutivas hechas por personajes femeninos (que no sean el mismo personaje). Si el programa encuentra al menos una conversación entre mujeres, considera que la película ha cumplido con la segunda condición del test. En este punto es donde pueden darse más discordancias con el análisis que haría un humano: el programa no es capaz de evaluar si dos mujeres hablan entre sí cuando hay alguna intervención en el medio. Por otro lado, el programa no es capaz de dividir el guión en escenas, con lo cual se dan casos en que si la última intervención de una escena y la primera intervención de la escena siguiente están hechas por personajes femeninos, el programa lo identificará como un diálogo. Como ejemplo paradigmático, nuestro programa estableció que la película *Magnolia* pasa el Test de Bechdel, habiendo un solo “diálogo” entre mujeres que no es acerca de un hombre: “ROSE:-bye CLAUDIA: -Hello”. Esta incapacidad del programa se debe a que el formato de los guiones no es lo suficientemente regular para poder identificar el comienzo o el final de las escenas. Si bien existen protocolos en relación a la estructura y formato que debieran tener los guiones de películas,⁹ en la práctica no existe un único formato. Por suerte, esto no constituye un problema para identificar los personajes y separarlos de las intervenciones, dado que a pesar de los distintos formatos algunas reglas generales se suelen cumplir.

Identificación de referencias a hombres. Para determinar si una película pasa la tercera condición, el programa recorre las conversaciones entre personajes femeninos buscando referencias masculinas, que identifica comparándolas con una lista generada de la siguiente manera: en primer lugar, existe una lista base que incluye nombres masculinos y referencias masculinas (him, husband, father, entre otros). En segundo lugar, cada vez que el programa identifica los personajes femeninos de un guión, añade a esta lista al resto de los personajes. Esta decisión puede traer algunos errores, pero hemos visto que aplicarla mejora significativamente la eficiencia del test, cuando se comparan los resultados obtenidos con los que arroja el sitio bechdeltest.com.

Validación del programa. Utilizamos dos corpus de guiones para poner a prueba el programa, ambos extraídos de The Internet Movie Script Database.¹⁰ El corpus mayor contiene 946 guiones (casi la totalidad de guiones en dicho sitio). El corpus menor contiene un subconjunto de guiones de películas del corpus mayor (176), aquellas cuyo resultado para el Test de Bechdel se encuentra en el sitio bechdeltest.com. Para encontrar dicha intersección se bajaron los nombres de todas las películas del sitio bechdeltest.com junto con su resultado en el test. El corpus menor nos permitió comparar los resultados obtenidos por nuestro programa en el transcurso de su desarrollo, de manera de ir modificando funciones que mejoraran su eficiencia. Además, del corpus menor se eliminaron algunas películas cuyos personajes tenían nombres que no iban a poder ser identificados por nuestro programa por el hecho de ser extraterrestres (por ejemplo, *Star Trek*, *Star Wars*).

Versión masculina del test. Si bien puede parecer evidente que los requerimientos del test no son muy exigentes, resulta adecuado leer los resultados en comparación con un test inverso, sobretodo si el análisis se hace de forma cuantitativa. La versión masculina del Test de Bechdel que desarrollamos es la siguiente: una película pasa el test si hay al menos dos hombres que en algún momento hablan entre sí acerca de algo que no es una mujer. De esta manera, podemos estar seguros que los resultados se deben a desigualdad de género y no artefactos de la herramienta utilizada.

Análisis por género de película. Como una prueba de concepto de una posible aplicación del test automático se estudió la proporción de películas que fallan el test en función de los distintos géneros cinematográficos. Este análisis fue realizado sobre el corpus mayor, para lo cual se asignó cada película a uno o varios géneros. Esta infor-

mación estaba disponible en el mismo sitio de donde los guiones fueron obtenidos.¹¹ Los géneros para los cuales se tenían menos de 100 películas fueron descartados del análisis. Los resultados obtenidos fueron luego comparados con un análisis independiente sobre la base de datos de bechdeltest.com, realizado con anterioridad.¹²

Resultados

Validación del programa. Utilizando el programa desarrollado aplicamos el Test de Bechdel al corpus menor de guiones, encontrando que un 45.61 % no pasa el test. Estos datos son similares a los arrojados por el sitio web, según el cual lo falla el 46.7% de las películas. Por lo tanto creemos que la herramienta desarrollada es robusta, si bien hay algunos parámetros que podrían mejorarse. Por otro lado estos resultados confirman el sesgo existente entre ambos géneros. Para entender a qué pueden deberse las diferencias existentes entre los resultados de ambas herramientas, desagregamos los mismos en función de la cantidad de condiciones que cumplen las películas (0, 1, 2 ó 3). La Tabla III A muestra las probabilidades de dar la respuesta *i* dado que la película clasifica como *j* según el test del sitio web. Si los resultados de nuestro programa coincidieran exactamente con los de la página web la tabla debería tener 1 en la diagonal y 0 fuera. Si bien no coinciden exactamente, se ve que los elementos de la diagonal son mayores a los elementos fuera de la diagonal. Las mayores diferencias parecen deberse a que nuestro programa asigna un “3” (es decir, considera que se han cumplido las 3 condiciones) en muchos casos en los que el sitio asigna un 1 ó un 2. Esto podría deberse a que nuestro programa identifica conversaciones entre mujeres que no lo son (por ejemplo, debido a cambios de escenas) y/o no es capaz de identificar algunas referencias masculinas.

	Res. Prog.	0	1	2	3
Res. Sitio					
0		0.65	0.18	0.06	0.12
1		0.07	0.37	0.23	0.33
2		0.03	0.03	0.47	0.47
3		0.05	0.06	0.11	0.78

Tabla IIIA. Probabilidades P (III) de que el programa de la respuesta *i* dado que la película clasifica como *j* en el sitio web.

Por otro lado, analizando algunas películas hemos visto que no es solo nuestro test el que falla. Se vio que la base de datos del sitio también tiene errores, pues con nuestro programa se identificaron guiones que claramente pasaban el test y que en el sitio clasificaban como que no lo pasaban. Por ejemplo, en la película *El Gran Pez* el único diálogo hallado entre mujeres fue -“Voy a lavar los platos”- “Yo te ayudo”. Según bechdeltest.com la película no pasa la tercera condición, si bien en el foro uno de los usuarios manifiesta su desacuerdo justamente haciendo referencia a este diálogo. Por otro lado, este es un buen ejemplo de lo aclarado respecto a la interpretación del test: el hecho de que una película en particular lo pase no implica que la participación femenina sea muy destacable. Al analizar el corpus de datos mayor, encontramos radicales diferencias en relación al número de películas que pasan el test. Si bien utilizando el corpus menor sabemos que hay cosas que se pueden mejorar, decidimos averiguar qué resultados se obtienen al correr el programa con un corpus más grande de guiones. Se encontró que un porcentaje mucho menor de películas pasaban el test, exactamente un 21.45 %. Esto podría deberse a varias causas. En primer lugar, el corpus menor, hecho con la intersección del corpus mayor con la base de datos del sitio, puede tener un cierto sesgo. Es decir, que las películas mayormente subidas por los usuarios del sitio sean una muestra no representativa del universo de guiones, con un mayor número de películas que sí pasan el test. Otra opción (no excluyente) es que el corpus mayor contenga un mayor porcentaje de películas que nuestro programa no es capaz de analizar correctamente. Esto concuerda con el hecho de que del corpus menor se eliminaron algunas películas cuyos personajes tenían nombres especialmente inventados. Dada la tendencia del test automático a asignar un “3” en casos donde el sitio web asignaría “1” o “2” (tendencia observada sobre el corpus menor), es llamativo el hecho de que al aumentar el tamaño la cantidad de películas que pasan el test disminuya tan abruptamente. Esta observación parecería apoyar la hipótesis de que las películas ingresadas en el sitio web están sesgadas.

Versión masculina del test. Utilizando un programa inverso al desarrollado para aplicar el Test de Bechdel, analizamos el corpus menor de guiones, obteniendo que un 87.64 % de los guiones pasan el test inverso.

Análisis por género de películas. Los guiones del corpus mayor fueron catalogados según su género y analizados por el test automático. Un análisis similar fue realizado previamente sobre la base de datos de bechdeltest.com.¹³ Ambos resultados se muestran en la Figura 2.

Solo 6 géneros cumplieron con la condición de tener más de 100 películas: Romance, Drama, Comedy, Thriller, Crime, Action. Se observa que el test automático ordena los géneros según la cantidad de películas que pasan en una forma consistente con los resultados del sitio web. Esto nos permite concluir que el test desarrollado es una poderosa herramienta para realizar estudios automáticos sobre grandes corpus de datos en función de parámetros tales como géneros cinematográficos.

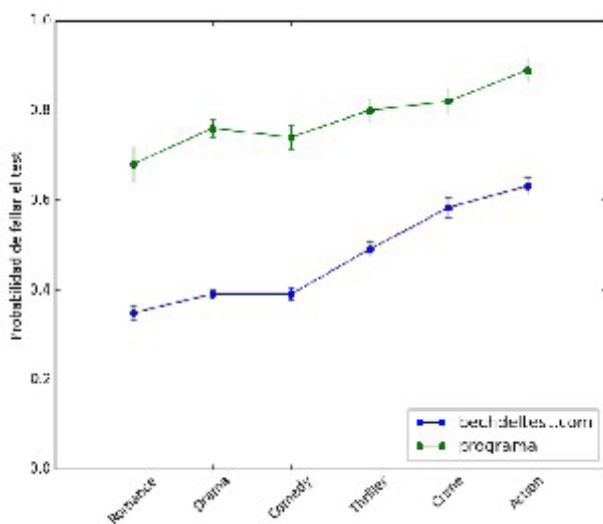


Figura 2. Probabilidad de fallar el Test de Bechdel en función del género cinematográfico. Resultados del test automático (verde) y del sitio web bechdeltest.com (azul).

Discusión

Desde la publicación original de la tira *The Rule* en 1985, el mayor intento por realizar estudios sistemáticos

empleando el Test de Bechdel ha sido el desarrollo del sitio web bechdeltest.com. Este sitio cuenta con una base de datos muy grande, que crece con el constante aporte de los usuarios. Asimismo, sobre esa base de datos se han realizado algunos análisis estadísticos que permiten ver la respuesta al test para películas de diferentes años, directores, género cinematográfico, etc. Sin embargo, hasta el momento no se había desarrollado ningún programa automático que fuera capaz de aplicar el test sobre guiones de películas, ni de analizar grandes corpus de guiones en función de los parámetros nombrados. La principal ventaja de un test automático es que permitiría seleccionar libremente la muestra de películas a analizar, siempre y cuando estén disponibles los guiones. En cambio, los resultados obtenidos en este mismo trabajo parecen indicar que los usuarios del sitio tienden mayormente a subir películas que pasan el test. La simpleza de las condiciones del test permite el desarrollo de un programa relativamente sencillo que da lugar a un análisis automático. Remarcamos que nuestro desarrollo es una primera aproximación muy simple al problema, y que quedan muchos desafíos por resolver. Algunos de estos desafíos son de naturaleza técnica o metodológica, como por ejemplo la separación de un guión en escenas (que permitiría identificar mejor los diálogos), mientras que otros son limitaciones inherentes al análisis de un texto. Por ejemplo hay información gestual o referencias visuales que el programa no es capaz de analizar, especialmente en el idioma inglés, en el que los adjetivos no tienen género. A pesar de estos problemas y la sencillez de nuestro método, los resultados parecen indicar que un test de Bechdel automático es posible. Más aún, fue posible utilizar el test desarrollado para ordenar distintos géneros cinematográficos en base a la proporción de películas que pasan el test.

Referencias

- 1 Bechdel, Alison. *Dykes to watch out for*. Firebrand Books, 1986.
- 2 Woolf, Virginia. "A Room of One's Own. 1929." New York: HBJ (1957).
- 3 <http://bechdeltest.com>
- 4 <http://bechdeltest.com/statistics/>
- 5 G. van Rossum, Python tutorial, Technical Report CS-R9526, Centrum voor Wiskunde en Informatica (CWI), Amsterdam, May 1995.
- 6 <http://www.crummy.com/software/BeautifulSoup/>
- 7 http://www.census.gov/genealogy/www/data/1990surnames/names_files.html
- 8 http://es.wikipedia.org/wiki/Guion_cinematogr%C3%A1fico
- 9 <http://www.imsdb.com/>
- 10 http://www.reddit.com/r/dataisbeautiful/comments/1hn113/ten_graphics_on_the_bechdel_test_oc

¹¹ <http://www.imsdb.com/>

¹² http://www.reddit.com/r/dataisbeautiful/comments/1hn1l3/ten_graphics_on_the_bechdel_test_oc

¹³ <http://bechdeltest.com>